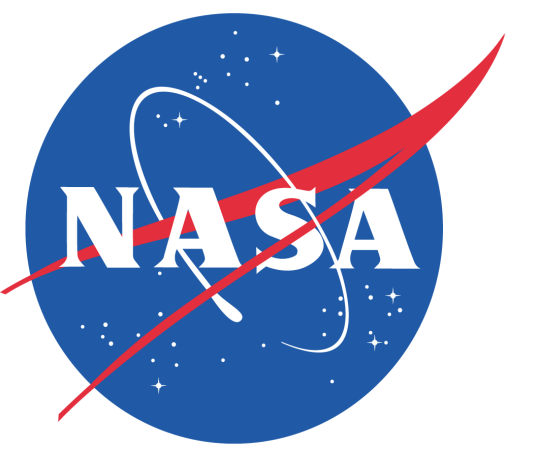


# Deriving Earth Science Data Analytics Requirements



AGU/15  
IN23C  
-1740

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

## Goal oriented Earth Science Data Analytics (ESDA)

reveal requirements for needed data

analytics tools/techniques

Steve Kempler<sup>1</sup>, ESIP ESDA Cluster\*

<sup>1</sup>NASA Goddard Earth Science Data and Information Services Center (GES DISC)  
Steven.J.Kempler@nasa.gov

### Motivation

How can we maximize the usability of large heterogeneous datasets to glean knowledge out of the data?

### Methodology

Categorize/Analyze ESDA use cases; derive data analytics requirements; associate tools/techniques; perform gap analysis

### Earth Science Data Analytics: Definition

The process of examining, preparing, reducing, and analyzing large amounts of spatial (multi-dimensional), temporal, or spectral data using a variety of data types to uncover patterns, correlations and other information, to better understand our Earth.

Data Preparation

Data Reduction

Data Analysis

### Earth Science Data Analytics: Goals

To validate data

To perform coarse data preparation

To intercompare datasets

To tease out information

To glean knowledge

To derive conclusions

To calibrate data

To assess data quality

To forecast/predict/model

To derive new analytics tools

### Earth Science Data Analytics: Initial Requirements

Ingest from various sources; Homogenize data; Visualization; Sampling; Gridding

Access large datasets; High speed processing; Subsetting, mining, machine learning

Homogenize data; Intercomparison statistics; Pattern recognition

Seek heterogeneous data relationships; Ingest from various sources; Image processing

Looking for Community input

Data exploration; Filter, mine, fuse, interpolate data; Manage custom code

Ingest from various sources; High speed processing; Math functions

Access large datasets; Assess erroneous data; Detect data anomalies

Data exploration; Neural networks; Math/Stat modeling; Near Real Time data

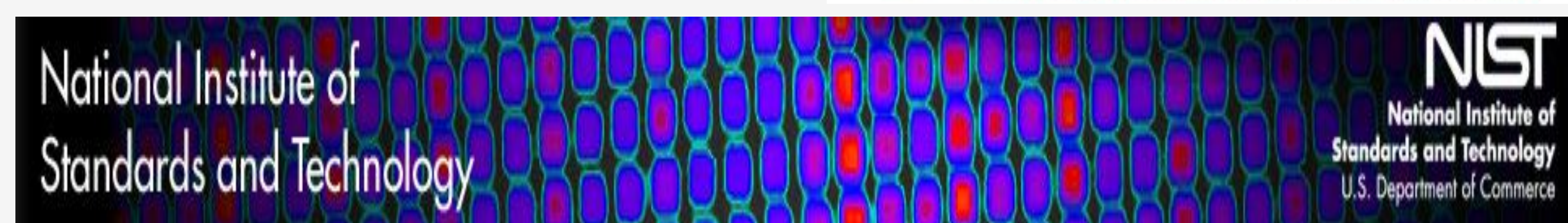
Access very large datasets; homogenize data; visualization

### Earth Science Data Analytics: Exemplary Tools, Techniques, Integrated Systems

Types of Analytics	Tools	Techniques	Integrated Systems
<ul style="list-style-type: none"><li>Data Preparation</li><li>Data Reduction</li><li>Data Analysis</li></ul>	<ul style="list-style-type: none"><li>R, SAS, Python, Java, C++</li><li>SPSS, MATLAB, Minitab</li><li>CPLEX, GAMS, Gauss</li><li>Tableau, Spotfire</li><li>VBA, Excel, MySQL</li><li>Javascript, Perl, PHP</li><li>Open Source Databases</li><li>PIO, NCL, Parallel NetCDF</li><li>AWS, Cloud Solutions, Hadoop</li><li>MPI, GIS, ROI-PAC, GDAL</li></ul>	<ul style="list-style-type: none"><li>Statistics functions</li><li>Machine Learning</li><li>Data Mining</li><li>Natural Language Processing</li><li>Linear/Non-linear Regression</li><li>Logical Regression</li><li>Time Series Models</li><li>Clustering</li><li>Decision Tree</li><li>Factor Analysis</li><li>Principal Component Analysis</li><li>Neural Networks</li><li>Bayesian Techniques</li><li>Text Analytics</li><li>Graph Analytics</li><li>Visual Analytics</li><li>Map Reduce</li></ul>	<ul style="list-style-type: none"><li>EarthServer (<a href="http://www.earthserver.eu">http://www.earthserver.eu</a>)</li><li>NASA Earth Exchange (<a href="https://nex.nasa.gov/nex/">https://nex.nasa.gov/nex/</a>)</li><li>EDEN (<a href="http://cda.ornl.gov/projects/eden/#">http://cda.ornl.gov/projects/eden/#</a>)</li><li>EARTHDATA (<a href="https://earthdata.nasa.gov">https://earthdata.nasa.gov</a>)</li><li>Giovanni (<a href="http://giovanni.gsfc.nasa.gov/giovanni/">http://giovanni.gsfc.nasa.gov/giovanni/</a>)</li></ul>

Compiled from: <http://practicalanalytics.co/predictive-analytics-101/> and <http://cda.ornl.gov/research.shtml>

### Earth Science Data Analytics: Enabling Organizations



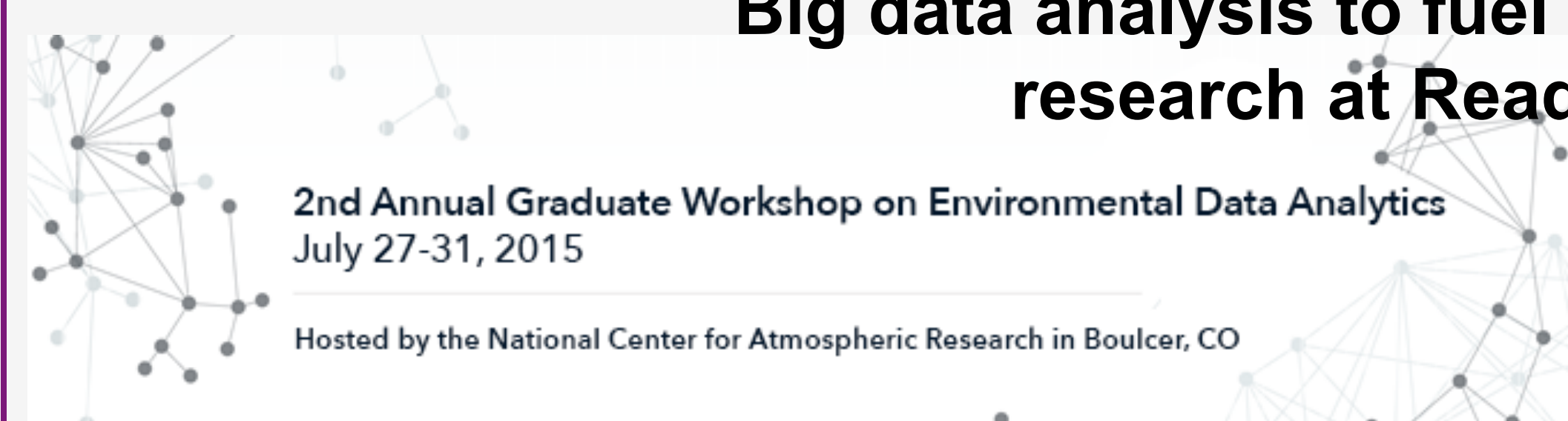
### The good news...

### Earth Science Data Analytics: Preparing for the Future

Central England NERC Training Alliance



Big data analysis to fuel environmental research at Reading University



... offering degrees in Data Science

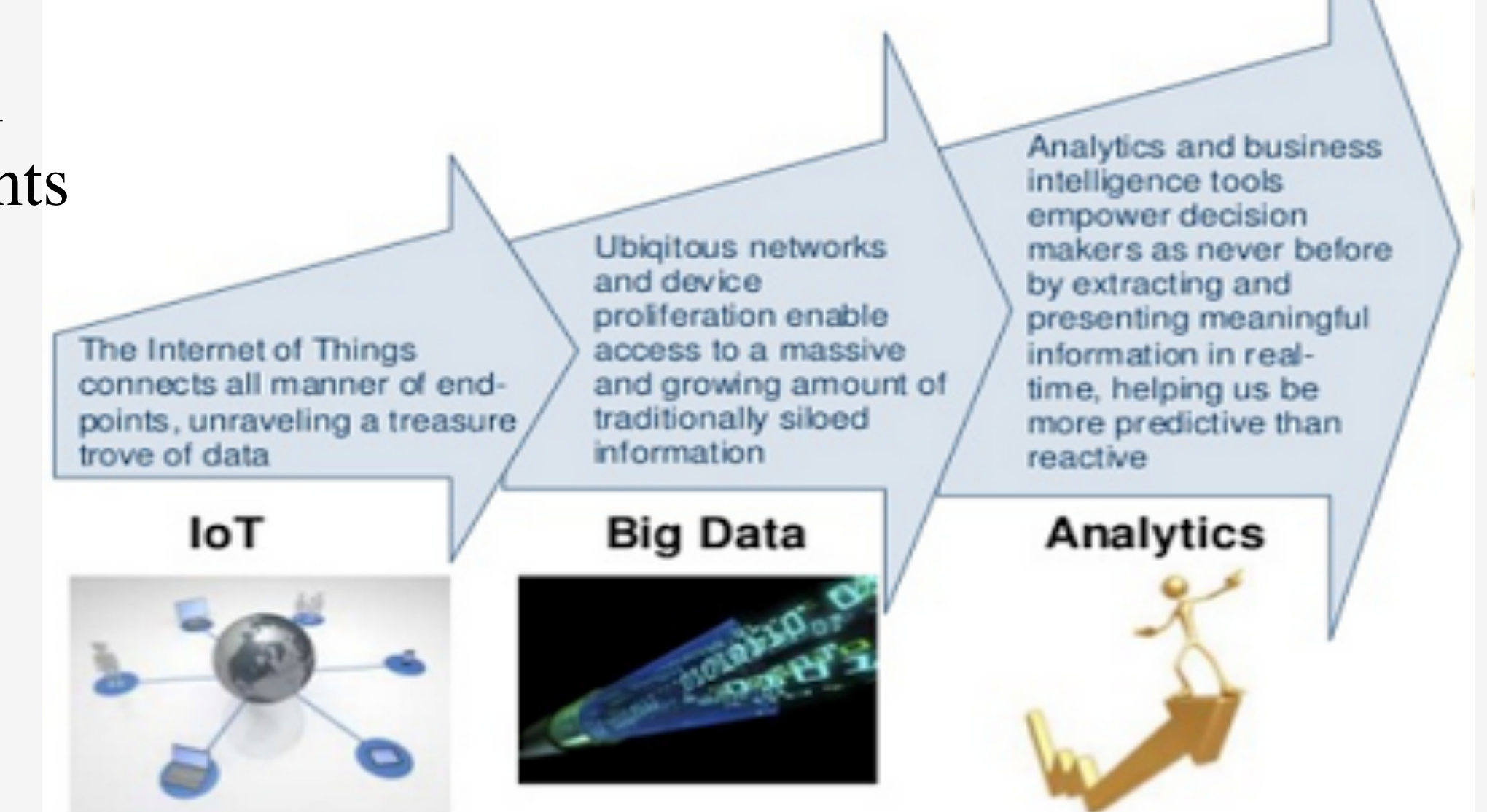
... summer school on Big Data Analytics

... online master's degree in data analytics

### Earth Science Data Analytics: Looking Ahead

• Complete Gap Analysis between ESDA requirements and current tools/technologies

• Continue to evolve tools/techniques to address growing scope of the 'Internet of Things'



<http://hashtaggers586.blogspot.com> (Source: <http://iot.sys-con.com/>)

\* Thanks to the work of the Earth Science Information Partners (ESIP) Federation, Earth Science Data Analytics (ESDA) Cluster